**SPRÅK8878 Linguistic Theory 5 (LingPhil)(at NTNU)**

**Studying language from observational data: inference, causality and limitations**

Instructor: Damián E. Blasi1,2

1: Department of Comparative Linguistics, University of Zürich, Switzerland

2: Department of Linguistic and Cultural Evolution, MPI for the Science of Human History,

Germany

**Description**

It is commonplace to note that the XXI century came with unprecedented amounts of data and

statistical methods relevant for the study of language. Data-oriented science doesn’t only

provide ancillary tools to traditional research on language, but it can also serve as a source of

new hypotheses, as a principled way of discovering complex patterns in data and as a safeguard

against some of the biases that affect qualitative research. On the other hand, inferences with

data come with their unique set of limitations and opportunities for misuse. In this course, I’ll

provide a brief introduction to data science, revisiting classic statistical topics as well as recent

methodological developments from the last decade. I’ll illustrate the concepts with case studies

taken from my work as well as from others, ranging linguistic typology, language acquisition,

historical linguistics and psycholinguistics.

**Structure of the course**

Session 1: The statistical turn in the language sciences

Session 2: A recap of some fundamental statistical ideas

Session 3: Causal inference with observational data

Session 4: Bayesian inference

Session 5: Inference from published studies: meta-analysis, bias and cumulative science

**Evaluation and credits**

Students will receive 5 ECTS for this course. The requirements are (a) active participation to the

lectures, (b) reading relevant references (from the general readings listed below, and as further

specified during the lectures), and (c) a 2000 word essay on a task determined by the lecturer

(Damián E. Blasi) jointly with the course organizer at LingPhil (Giosuè Baggio).

**General readings**

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian

data analysis (Vol. 2). Boca Raton, FL: CRC press.

Pearl, J. (2009). Causal inference in statistics: An overview. Statistics surveys, 3, 96-146.

Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models.

Cambridge university press.

Heinze-Deml, C., Maathuis, M. H., & Meinshausen, N. (2017). Causal structure learning. Annual

Review of Statistics and Its Application, (0).

McElreath, R. (2014). Statistical Rethinking. CRC Press.